Chemistry Add-in for Word User's Guide

Version 1.0 – January 28, 2011

Abstract

This document describes how to use the Chemistry Add-in for Word, an add-in for Microsoft Word that provides a simple and flexible way to include chemical information in a Word document.

Note:

- Most resources discussed in this paper are provided with the Chemistry Add-in package. For a complete list of documents and software discussed, see "Resources" at the end of this document.
- For Chemistry Add-in updates and software availability news, see http://chem4word.codeplex.com or http://research.microsoft.com/chem4word/
- For general news and discussion about the Chemistry Add-in for Word, please join our FaceBook page at http://www.facebook.com/home.php?sk=group_186300551397797

***Disclaimer**: This document is provided "as-is". Information and views expressed in this document, including URL and other Internet Web site references, may change without notice. You bear the risk of using it.*

This document does not provide you with any legal rights to any intellectual property in any Microsoft product. You may copy and use this document for your internal, reference purposes.

© 2011 Outercurve Foundation. All rights reserved.

Microsoft and Windows are trademarks of the Microsoft group of companies. All other trademarks are property of their respective owners.

Contents

Introduction 3
Getting Started 4
Prerequisites 4
System Requirements 4
Installation 4
UI Overview 5
How to Create and Manage Chemistry Zones 6
Specify a Chemistry Zone's Representation 7
Change a Zone's Labels 8
Insert a Zone from the Chemistry Gallery 9
Insert a Chemistry Zone from the Web 10
Manage the Zones in the Chemistry Gallery 11
How to Manage Chemistry Zones with the Chemistry Navigator 11
How to Edit 2-D Structures 13
The Controls of the 2-D Editor 14
Select Atoms for Editing 15
Edit Individual Atoms 16
Delete Selection 17
Change Atom Type 17
Set Label/Remove Label 19
+e/-e 19
Undo/Redo 19
$+H^+/-H^+$ 19
$+H^\bullet/-H^\bullet$ 19
Set Isotope/Remove Isotope 19
Change the Bonding 19
Modify the Diagram's Structure 20
Flip the Diagram 20
Move Atoms or Groups 20
Rotate Atoms or Groups 21
Delete Atoms or Groups 22
Resources 23
Appendix 1: CML Basics 24
CML Data Storage 24
Appemdix 2: How CML Data is Stored in a Document 25

© 2011 Outercurve Foundation. All rights reserved.

Introduction

The Chemistry Add-in provides a simple and flexible way to include chemical information in a Word document.

With the Chemistry Add-in, you can:

- **Create inline "chemistry zones" to represent chemical data.**

  Chemistry zones are controls that contain information about a molecule and display the information in a variety of ways. The underlying data is stored as Chemical Markup Language (CML), a widely used XML schema for representing chemical data. The data typically includes trivial and International Union of Pure and Applied Chemistry (IUPAC) names, the concise formula, and data for a 2-D structure.

- **Display chemical information in a variety of ways.**

  A chemistry zone can display any representation that is supported by the underlying CML data. With a few clicks, you can switch from the molecule's trivial name to its concise formula to its 2-D representation.

- **Display print-quality 2-D chemical structures.**

  Chemistry zones can represent molecules by displaying a 2-D structure diagram. The Chemistry Add-in also includes an editor that enables you to modify the structure. The diagram is inserted into the document as a PNG image, so that others can view it, whether or not they have installed the Chemistry Add-in on their system. You can also publish a document authored using the Chemistry Add-in as a PDF file.

- **Accept chemical data in a variety of formats.**

  You can create a "chemistry zone" by typing a simple common name such as "water", and then using the Chemistry Add-in to convert it to your preferred representation.

- **Import CML files from online web services**

  Using the **Load From** option in the ribbon, you can look up existing molecular structures from the NCBI's PubChem (http://pubchem.ncbi.nlm.nih.gov/) or the Unilever Centre's OPSIN (http://opsin.ch.cam.ac.uk/) databases.

- **Handle most molecules.**

  The Chemistry Add-in can handle any molecule that has appropriate CML data. The Chemistry Add-in includes CML data for several hundred common molecules. For additional molecules, you simply create or obtain a CML file and import it into a document.

- **Store and expose chemical information in a semantically rich manner.**

  The Chemistry Add-in supports publishing and data-mining scenarios for authors, readers, publishers, and other vendors across the chemical information community.

This paper describes how to use the Chemistry Add-in to include chemical information in a Word document.

© 2011 Outercurve Foundation. All rights reserved.

Getting Started

This section describes how to get started with the Chemistry Add-in.

Prerequisites

You should have a basic understanding of:

- Microsoft Word
- Chemical nomenclature and diagrams

An understanding of CML is helpful but not required.

System Requirements

Hardware Requirements

Any computer that can run Office 2007 or Office 2010.

Software Requirements

Your computer must have the following software:

- Any version of Windows that can run Office 2007 or Office 2010, which includes Windows XP SP3 and later versions of Windows.
- Word 2007 or Word 2010

For details on Office requirements, see the Resources section at the end of this document.

Installation

The Chemistry Add-in is packaged in a file named ChemistryAddinforWordv1.zip that contains the following files:

- Setup.exe
- Chem4Word.Setup.msi

To Install the Chemistry Add-in

1. Close all Word documents.
2. Copy the Chemistry Add-in .zip file to your hard drive.
3. Extract the contents of the .zip file to a folder.
4. Navigate to that folder and run Setup.exe, which runs the standard MSI installation wizard.
5. Use the installation wizard to install the Chemistry Add-in.

To verify the installation, launch Word. The ribbon should now include a Chemistry tab, as shown in Figure 1.

© 2011 Outercurve Foundation. All rights reserved.

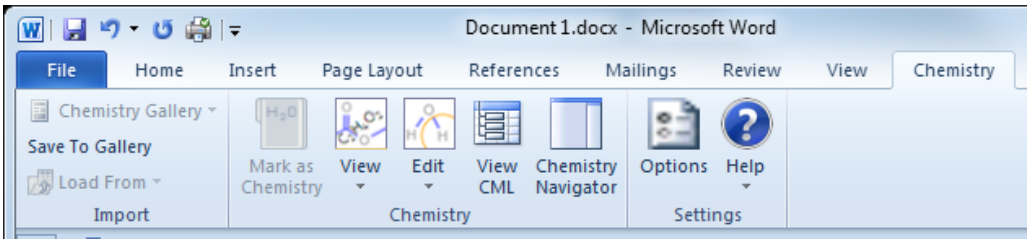


Figure 1. Word ribbon with Chemistry tab

Tip: The rest of the document will be easier to follow if you first install the Chemistry Add-in.

UI Overview

The Chemistry tab contains the primary user interface for the Chemistry Add-in, as shown in Figure 2.

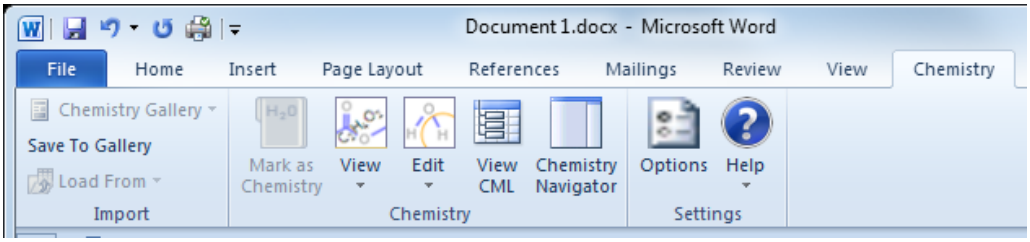


Figure 2. Chemistry tab

The tab has three command groups:

Import

Import and save chemical data:

- **Chemistry Gallery** displays a gallery of chemistry zones that can be inserted into a document.
- **Save To Gallery** saves a chemistry zone to the Chemistry Gallery for easy re-use in later documents.
- **Load From** imports a CML data file and adds it to the document

Chemistry

Manage chemistry zones:

- **Mark as Chemistry** marks selected text as a chemistry zone.
- **View** enables you to select which representation of a molecule to display.
- **Edit** displays the **Edit 2D** and **Edit Labels** options in order to modify the 2-D structure or the text associated with a zone.
- **Chemistry Navigator** displays a panel that contains all the chemistry zones in the current document.

Settings

Manage the chemistry-related aspects of the document:

- **Options** enables you to configure the Chemistry Add-in defaults.
- **Help** displays the **Help** and **Check for updates** commands.

© 2011 Outercurve Foundation. All rights reserved.

How to Create and Manage Chemistry Zones

A chemistry zone displays data for a particular molecule, based on the molecule's CML data. There are two kinds of chemistry zones: *textual* and *2D*. Textual chemistry zones are based on equation zones and have a similar UI. The 2D zones use a picture content control. This section shows how to use the commands from the **Chemistry** group to create and manage chemistry zones.

At its most basic level, a chemistry zone is simply a control that contains a selected block of text.

To create a chemistry zone from text

1. Select the word or phrase that you want to include in the zone and right-click it.
2. Click the **Mark as Chemistry** button and the corresponding text on the fly-out menu to create a chemistry zone for the selected text, as shown in Figure 3.

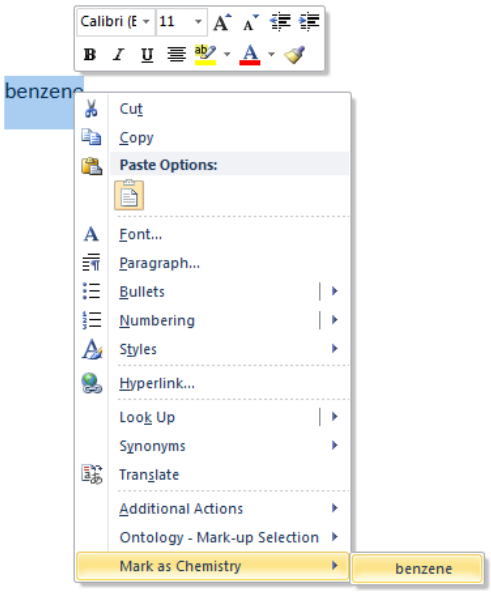


Figure 3. Creating a chemistry zone.

If you hover over a chemistry zone, Word highlights it. If you click the highlighted text, Word displays the zone UI, with a Chemistry label to identify it as a chemistry zone. Unlike Math equation zones, the chemistry zone UI does not display a dropdown menu. It simply indicates the zone's type and extent.

Each chemistry zone is based on CML data that is included in the document. Marking text as a chemistry zone automatically creates a CML data file which only contains the selected text as a chemical name of unknown type.

© 2011 Outercurve Foundation. All rights reserved.

Specify a Chemistry Zone's Representation

When you click **Mark as Chemistry**, the add-in adds the molecule's CML data to the document, which includes:

- One or more text labels (a molecule's CML data usually includes the IUPAC name and usually one or more trivial names).
- The concise formula (e.g. C_6H_6).
- 2-D structural data.

The benzene CML data contains two labels—the IUPAC name (benzene) and a trivial name (phenyl hydride)—the concise formula, and a set of 2-D data.

The CML infrastructure recognizes all of benzene's names and its concise formula, so you can use any of them to enter benzene into your document and convert it to a chemistry zone. After you have converted an instance of "benzene"—or one of its equivalents—into a chemistry zone, you can easily change the representation.

To change a chemistry zone's representation

1. Click the zone to select it.
2. Click the **Chemistry** group's **View** command, which will display a drop-down list of the possible display options supported by add-in based on the molecule's CML data. Figure 4 shows the dialog box for benzene.
3. Select the preferred representation from the list.

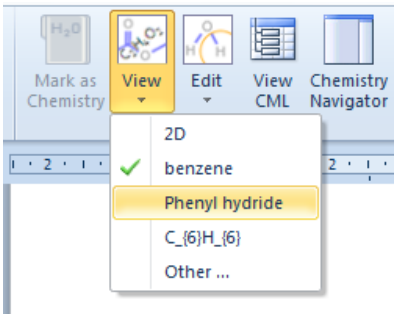


Figure 4. Select View drop-down

The "2D" option displays a structural diagram, as shown in Figure 5.

Figure 5. Benzene's 2-D representation.

© 2011 Outercurve Foundation. All rights reserved.

Change a Zone's Labels

A molecule's CML data can include any number of labels. For example, the CML data for benzene includes two labels, but water includes seven. You can modify a zone's labels with the **Edit Labels** command.

To modify a zone's labels

1. Select the zone.
2. Click **Edit**, and then **Edit Labels**.

   Figure 6 shows the **Edit Labels** dialog box for water.

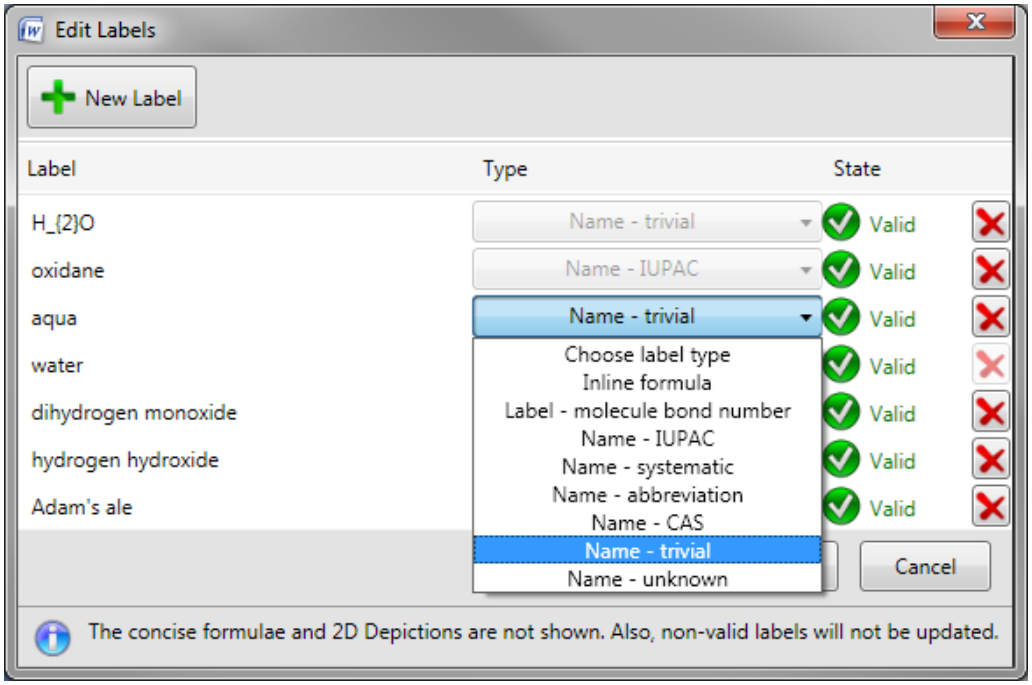


Figure 6. Edit Labels dialog box.

The Chemistry Add-in enables you to add or remove labels (if they are not being used in the document), or modify their properties.

To add a label

1. Click **New Label**, which displays an empty label field below the last label on the current list.
2. Type the new label's name in the empty box.
3. Select the label type from the **Choose label type** dropdown to the right of the name. A label can have one of the types shown in Figure 6.
4. Click **OK** to add the new label to the zone's CML.

You can remove labels or modify their properties.

- To remove a label, click the X at the label's right edge.
- To modify a label's text, click the text box and edit the text.

© 2011 Outercurve Foundation. All rights reserved.

- To change the label type, select a new type from the dropdown.

When you are finished, click **OK** to accept the changes.

The changes apply only to the selected zone's CML data—which is stored in the document—and to any linked zones. It does not change the CML file used to create the zone. If you use the procedure in the preceding section to create a new water zone, it will have a default set of labels. Linked zones are discussed later in this paper.

Insert a Zone from the Chemistry Gallery

An alternative way to add chemistry zones to a document is the **Chemistry Gallery**, which contains a collection of zones that can be inserted directly into a document.

Figure 7 shows the first few items in the **Chemistry Gallery** that is included with the the Chemistry Add-in package.

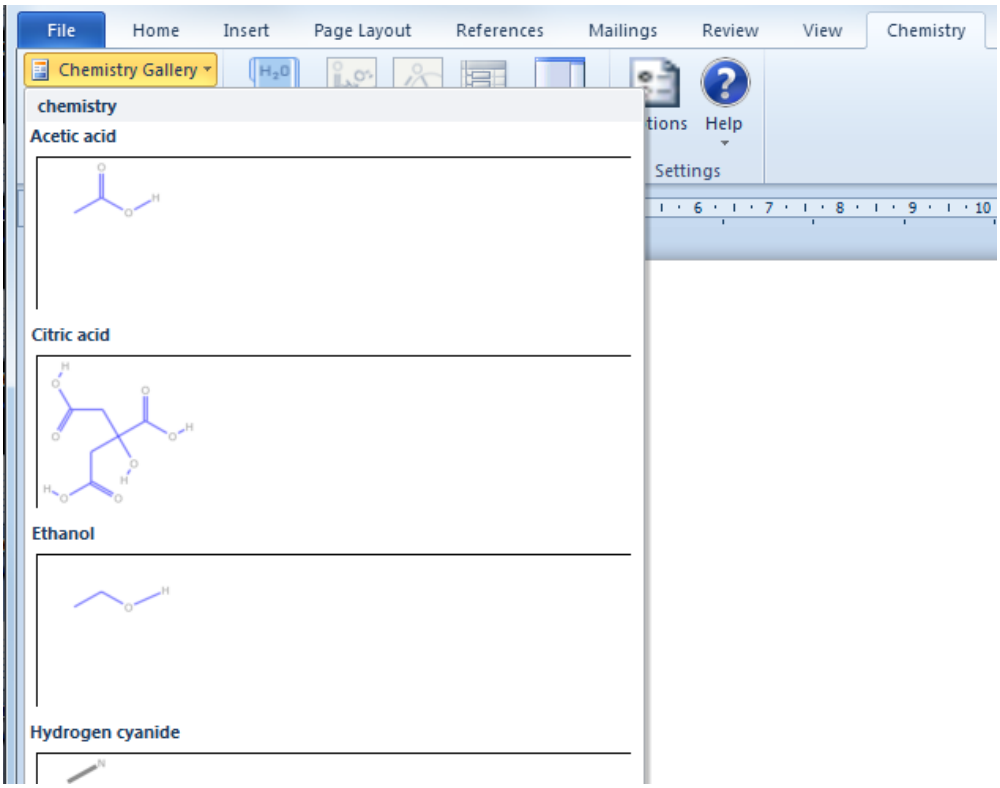


Figure 7. Chemistry Gallery

To insert a chemistry zone into a document from the Chemistry Gallery

1. Place the cursor at the appropriate place in the document.
2. Click **Chemistry Gallery** to display the gallery.
3. Click the appropriate molecule to insert the associated chemistry zone into the document.
4. Change the zone's representation to the appropriate representation.

© 2011 Outercurve Foundation. All rights reserved.

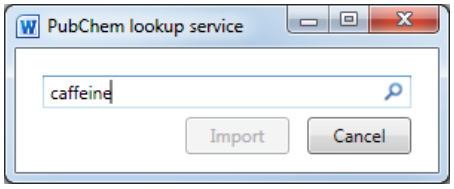
Insert a Chemistry Zone from the Web

Although the add-in ships with several hundred CML files, you can also search the following online databases and import chemical structures directly:

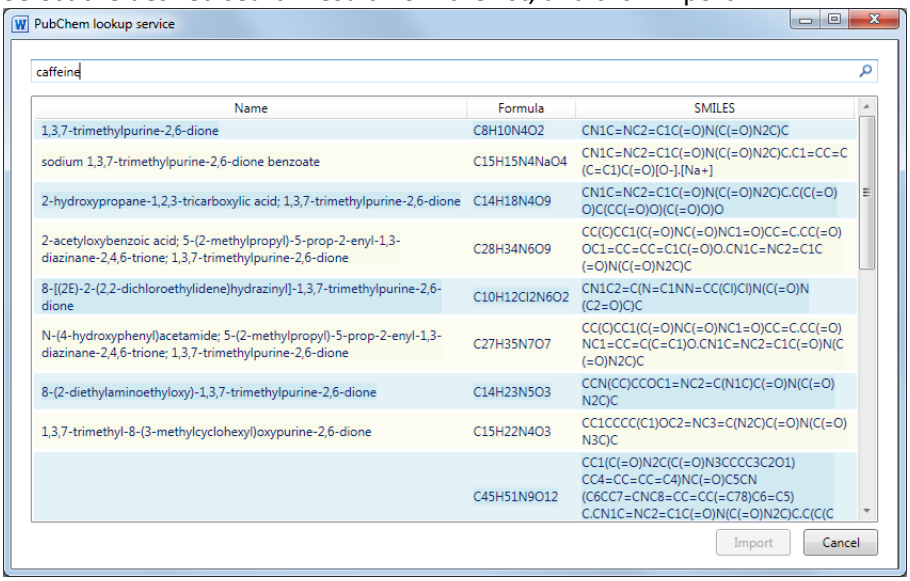
- PubChem
  - http://pubchem.ncbi.nlm.nih.gov/
  - PubChem is a chemical structures database provided by the National Center for Biotechnology Information (NCBI), part of the U.S. National Library of Medicine
- OPSIN
  - http://opsin.ch.cam.ac.uk/
  - OPSIN is a database maintained by the University of Cambridge whch converts IUPAC compound names to semantic chemical information including a connection table.

To insert a chemistry zone from the web

1. Place the cursor at the appropriate place in the document.
2. Click **Load From** in the Import section of the ribbon
3. Select **PubChem**



4. Type in you search term, and click the magnifying glass icon
5. Select the desired search result from the list, and click Import.



© 2011 Outercurve Foundation. All rights reserved.

Manage the Zones in the Chemistry Gallery

You can customize the **Chemistry Gallery** for your particular project or group by adding new zones in either of two ways:

- Insert a zone from your document.
- Import a CML file, to add zones that are not already included with the Chemistry Add-in.

  This option is discussed in "How to Import Chemical Data," later in this document.

To add a zone

1. Select the appropriate 2D chemistry zone.
2. Click **Save to Gallery**.
3. Specify a name for the zone and click **OK** to add the zone to the gallery.

It is only possible to add 2D zones to the gallery

Each item in the **Chemistry Gallery** represents a particular chemical zone, not necessarily a unique molecule. For example, if you modify a benzene zone by adding a new label or changing the 2-D representation, you can add that zone to the gallery as a separate item.

To move or delete a zone

1. Open the **Chemistry Gallery** and right click a zone to display the **Chemistry Gallery** popup.
2. Click **Organize and Delete** to display the **Building Blocks Organizer** dialog box.
3. Use the dialog box to modify the zone order, or delete zones.

   The **Chemistry Gallery** zones are in the **Custom 5** gallery.

How to Manage Chemistry Zones with the Chemistry Navigator

The **Chemistry Navigator** helps you manage a document's chemistry zones. To open the manager, click the **Chemistry Navigator** command, which opens the **Chemistry Navigator** panel on the right side of the document.

The **Chemistry Navigator** panel shows every chemistry zone in the document in document order: left to right, top to bottom, as shown in Figure 8.

© 2011 Outercurve Foundation. All rights reserved.

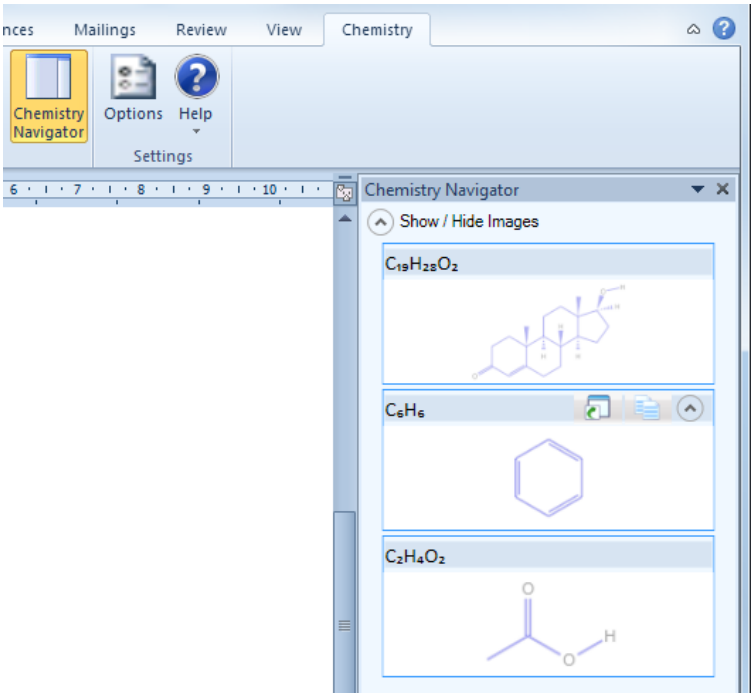


Figure 8. Chemistry Navigator panel

The items are represented in the panel by the preferred navigator depiction, followed by the 2-D representation where possible.

There are two zone types:

- ***Unlinked zones***: independent entities that have their own CML data.

  You can change an unlinked zone's representation or modify its labels or 2-D structure without affecting any other zone, even those that represent the same molecule.

- ***Linked zones***: a collection of zones that are based on the same stored CML data.

  If, for example, you change a particular linked zone's representation or modify its 2-D structure, you also change every other zone in the collection.

You can use **Chemistry Navigator** to insert linked or unlinked zones by using the buttons to the right of the appropriate concise formula, as shown in the second instance of C6H6 in Figure 8.

To insert a linked zone

1. Place the cursor at the appropriate place in the document.
2. Click the link icon (left) to insert the zone.
3. Specify the preferred representation in the document from the drop-down.

The new linked zone appears in the **Chemistry Navigator** in the appropriate document position.

© 2011 Outercurve Foundation. All rights reserved.

Note: You must create the first instance of a linked zone by using the techniques discussed earlier in this paper. Then you can use Chemistry Navigator to insert additional instances.

To insert an unlinked zone

1. Place the cursor at the appropriate place in the document.
2. Click the unlinked icon (right) to insert the zone.
3. Specify the preferred representation in the document from the drop-down.

The new unlinked zone appears in the **Chemistry Navigator** in the appropriate document position.

How to Edit 2-D Structures

The CML files can contain structural information such as the x-y position of the atoms. When you specify the 2-D representation to be used in the document, the Chemistry Add-in uses this data to generate and display a 2-D diagram. The Chemistry Add-in also enables you to modify a zone's 2-D representation, typically for one of the following reasons:

- To modify the positions of the atoms to make the diagram easier to read.
- To modify the composition of the molecule by replacing one or more atoms or by modifying bonds.
- To add labels to atoms.

Note: The current 2D editor can modify the geometry or composition of existing structures. However, its ability to add to the structure is limited, and you cannot create a structure from scratch.

To edit a zone's 2-D structure

1. Select a chemistry zone.
2. Click **Edit 2D** (available on the lower portion of the **Edit** command on the ribbon), which opens the 2-D editor in a separate window. The 2-D editor is only available as an option when the zone depicted in the document has an associated 2D structure.

   Figure 9 shows the 2-D editor, with a testosterone molecule. The window is modal, so you must close it before doing any further work in the document.
3. Edit the diagram, as described in the following sections.

© 2011 Outercurve Foundation. All rights reserved.

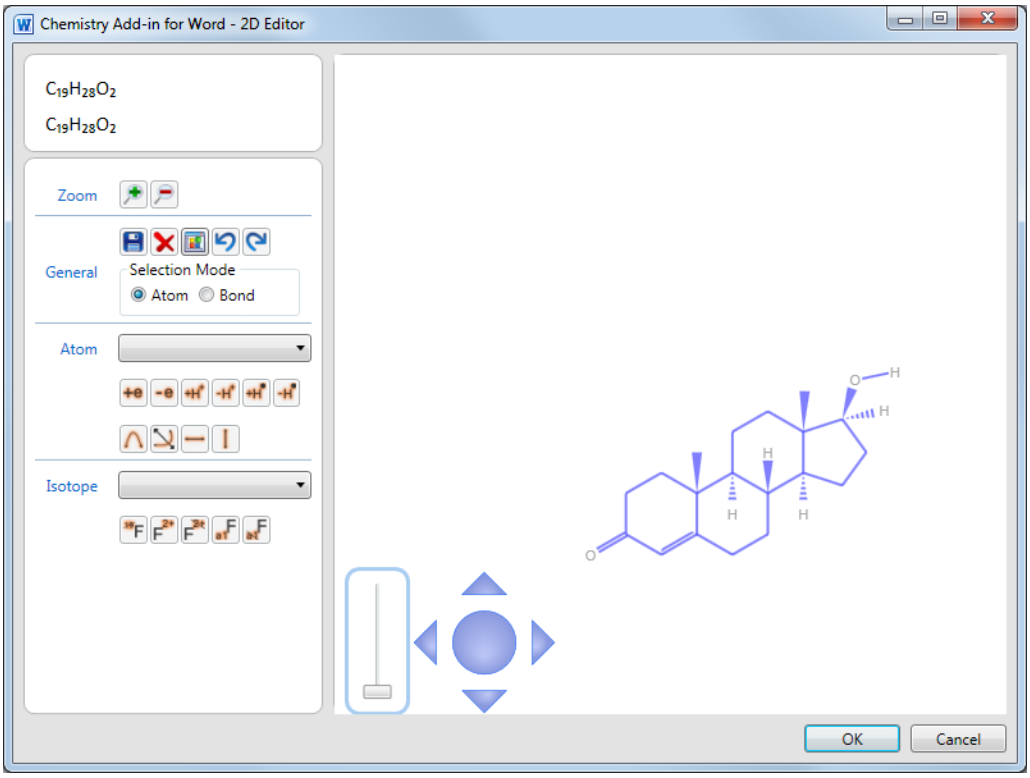


Figure 9. 2-D editor

When you click **OK** to close the editor, the zone's CML data is updated and the new diagram is displayed in the zone, as well as in any linked zones.

Important: When you modify a molecule, you will be prompted to delete and/or modify the labels. You may also use **Edit Labels** to assign new labels to the modified molecule.

The diagram shown in the document is a PNG image, which is generated from the CML data and inserted into the document. Others can thus read documents authored using the Chemistry Add-in, even with earlier versions of Word such as Word 2003. If you view a chemistry document with a version of Word that lacks the Chemistry Add-in, Word simply displays the image and ignores the embedded CML data.

The Controls of the 2-D Editor

The 2-D editor includes two controls for manipulating a diagram: a *zoom slider* and a *pan control*.

- **Zoom slider**: Click and drag the slider to zoom in and out of the diagram.
- **Pan control**: Click one of the four points of the pan control to pan the diagram in that direction. Click and drag the center of the pan control to pan the diagram in that direction.

© 2011 Outercurve Foundation. All rights reserved.

Select Atoms for Editing

Before you start editing, you must first select the atom or group of atoms that you want to work with.

To select single atoms

1. Position the cursor over the appropriate atom.

   The editor displays an orange circle when the cursor is properly positioned.

2. Clickthe orange circle to select the atom.

   The orange circle will change to a filled orange circle, and then to a filled blue circle when you move the cursor away.

3. To select multiple atoms, hold the CTRL key down and repeat steps 1 and 2.

   If you do not hold the CTRL key down, repeating steps one and two selects a different atom. CTRL Click on an atom which is already selected will deselect it.

Figure 10 shows testosterone with two atoms selected.

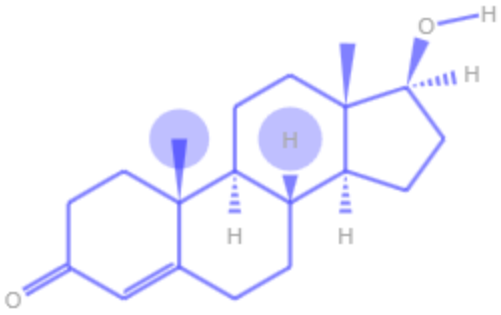


Figure 10. Selecting individual atoms

The editor supports two approaches to selecting multiple atoms in one operation, *marquee-select* and *gesture*.

To use marquee-select

1. Position the cursor just outside the “corner” of the group of atoms that you want to select.

   This approach selects all the atoms within a rectangle. This step defines one corner of the rectangle.

2. Click and hold the left button and drag the cursor so that the rectangle includes the atoms that belong in the group.

   Marquee-select is based on the positions of the atoms themselves, not the associated labels.

3. Release the left button.

The atoms in the group are indicated by closed blue circles, as shown in Figure 11.

© 2011 Outercurve Foundation. All rights reserved.

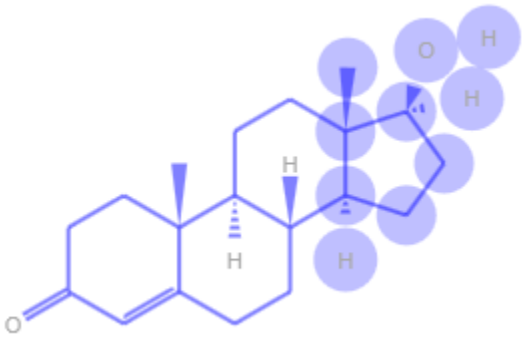


Figure 11. Marquee-select

Gesture selects multiple atoms "downstream" from a particular bond.

To use gesture to select a group of atoms

1. Position the cursor on the bond.
2. Click and hold the left button.
3. Move the cursor in the direction which you wish the selection to be. Release the mouse to select all the downstream atoms in this direction.

Tip: A green arrow appears to show which way you are going to be selecting. Moving back to the center of the bond will make the arrow disappear and nothing will be selected on release.

The hydroxyl group in Figure 12 was selected by gesturing along the carbon-oxygen bond.

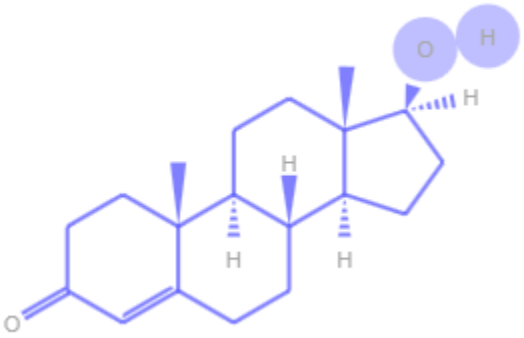


Figure 12. Gesture selection

Tip: It is sometimes convenient to move the molecule to a different location in the editor. To do so, select the entire molecule and drag it to the new location.

Edit Individual Atoms

If you hover over a selected atom, the editor displays the popup shown in Figure 13, which enables you to modify the atom in various ways. Hover over the icons to see the names of the commands.

© 2011 Outercurve Foundation. All rights reserved.

The **Rotate** and **Flip** commands are disabled, because they apply only to groups of atoms.

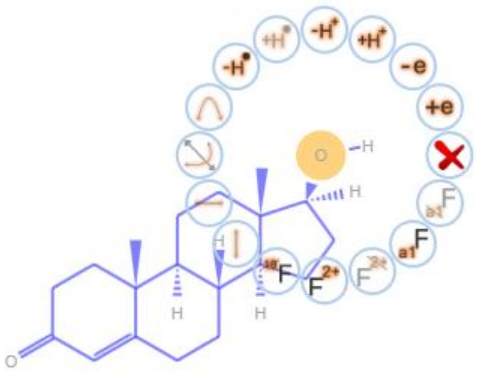


Figure 13. Editing popup

To edit the selected atom's properties, click the appropriate command, which typically displays a popup from which you can select alternatives. The remainder of this section discusses how to use the various commands to edit the properties of an atom.

Important: The 2-D editor attempts to provide guidance on the available chemically valid alternatives, but it does not enforce that guidance. For example, the **Set Isotope** popup lists likely isotope numbers for oxygen, but you can still choose to set the isotope number to 42.

Delete Selection

Delete Selection removes the atom from the diagram, and attempts to remove dependent atoms as well. For example, if you delete the highlighted O atom in Figure 13, the editor also deletes the associated H atom. Deleting any atom will also break all the bonds to this atom. The resultant unfilled valencies will be filled by adding the correct number of hydrogen atoms except where this would cause H_2 to be created.

Change Atom Type

This command changes the atom assigned to the selected location in the diagram. When you click **Change Atom Type**, the editor displays a dropdown list that enables you to select an alternative atom or group.

© 2011 Outercurve Foundation. All rights reserved.

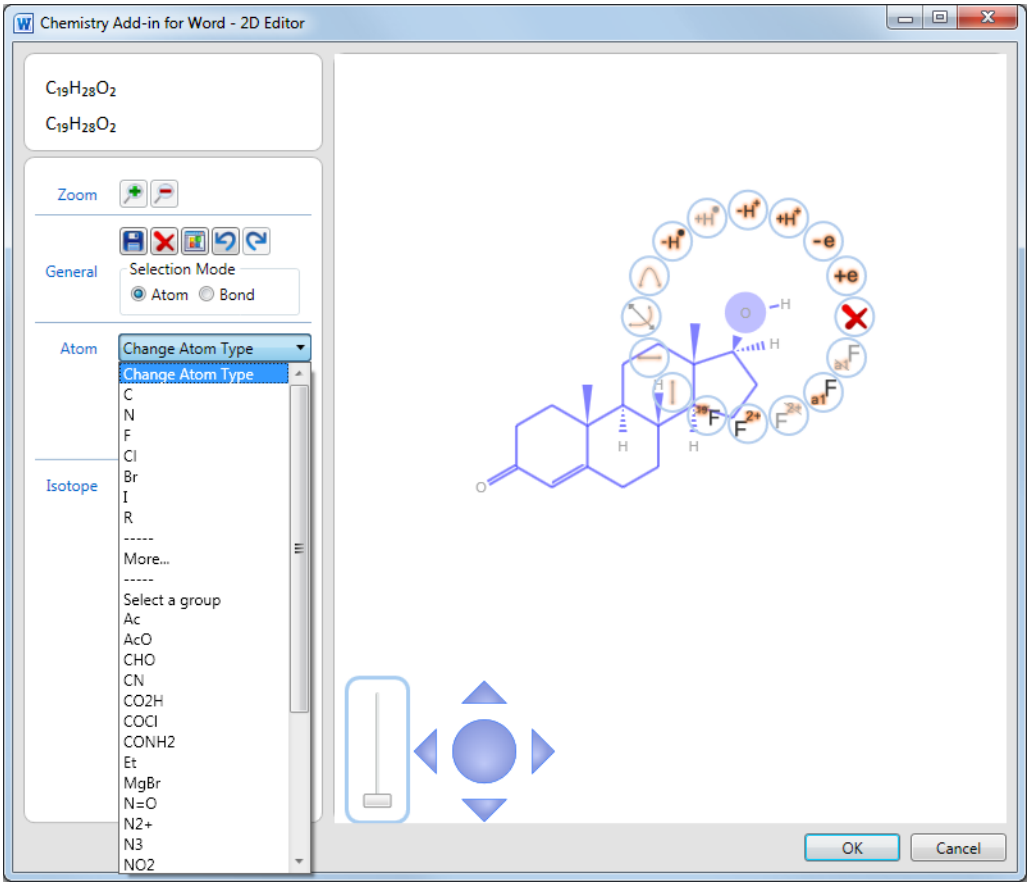


Figure 14. Change Atom Type drop-down menu

The first part of the list shows individual atom substitutions and the second a selected set of groups that can replace oxygen in the current chemical environment in a chemically sensible way.The editor adds or removes attached H atoms as required. However, you can insert any atom by clicking **More**. This item displays a periodic table—shown in Figure 15—from which you can select any atom by clicking it.

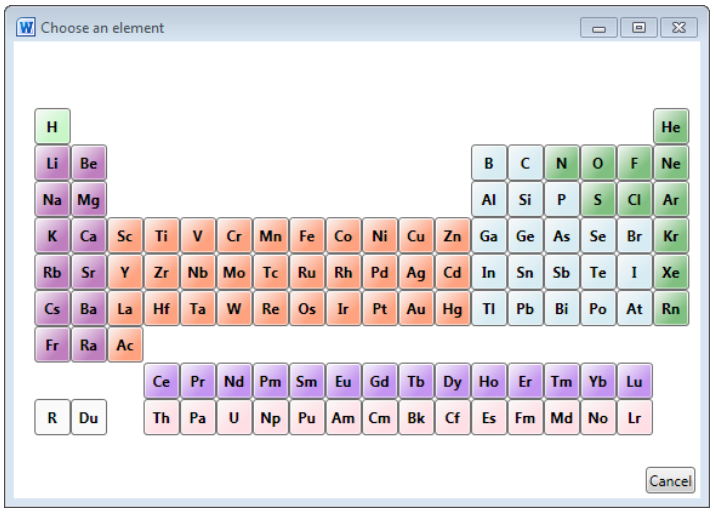

Choose an element
H
He
Li Be
B C N O F Ne
Na Mg
Al Si P S Cl Ar
K Ca Sc Ti V Cr Mn Fe Co Ni Cu Zn Ga Ge As Se Br Kr
Rb Sr Y Zr Nb Mo Tc Ru Rh Pd Ag Cd In Sn Sb Te I Xe
Cs Ba La Hf Ta W Re Os Ir Pt Au Hg Tl Pb Bi Po At Rn
Fr Ra Ac
Ce Pr Nd Pm Sm Eu Gd Tb Dy Ho Er Tm Yb Lu
R Du
Th Pa U Np Pu Am Cm Bk Cf Es Fm Md No Lr
Cancel

Figure 15. Periodic Table

© 2011 Outercurve Foundation. All rights reserved.

Set Label/Remove Label

These commands enable you to modify an atom's label by assigning an arbitrary string, which is displayed to the lower left of the label. Figure 16 shows the O atom from the previous examples, with an added "Test" label.

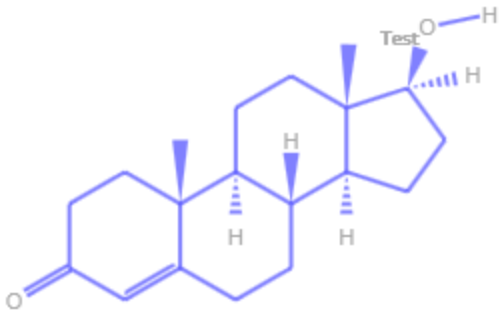


Figure 16. Adding a label to an atom.

+e/-e

These two commands add or remove electrons from an atom, and display the new state to the upper right of the atom's label.

Undo/Redo

These commands undo or redo your changes, beginning with the most recent.

$+H^+/-H^+$

These commands add or remove a proton from the selected atom.

$+H^\bullet/-H^\bullet$

These commands add or remove a hydrogen atom from the selected atom.

Set Isotope/Remove Isotope

Use the Set Isotope dropdown menu to select an isotope number. It is displayed at the upper left of an atom's label.

Change the Bonding

You can add one or more bonds to an existing bond, and change the bond group.

To change the bonding

1. Right-click anywhere in the diagram and select **Bond Mode**.
2. Select a bond in the diagram and the bond popup is displayed.
3. Use the **Select bond order** dropdown to add one or more bonds to the selected bond.
4. Use the **Select bond group** dropdown to select a bond group for the selected bond.

© 2011 Outercurve Foundation. All rights reserved.

5. Use the **Select bond stereo** dropdown to specify the stereo chemistry about an atom.

Figure 17 shows the upper right part of the testosterone molecule after a double bond has been added into the right-most ring. Notice that the editor removed two hydrogen atoms, to compensate.

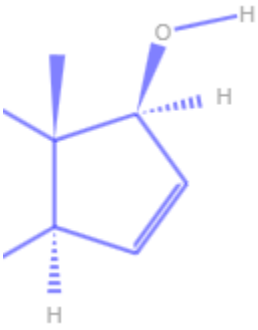


Figure 17. Add a bond

Modify the Diagram's Structure

You can modify the diagram's geometry in several ways, typically to improve readability.

Flip the Diagram

You can use the Flip commandsto flip the diagram.

- **FlipVertical:** rotates the diagram about a vertical axis.
- **Flip Horizontal:** rotates the diagram about a horizontal axis.
- **Flip About Bond**: rotates the selected atoms about the vector of the selected external bond.

To flip the diagram, select any atom and click the appropriate button. Figure 18 shows the default testosterone diagram, followed by how it is changed by **Flip Vertical** and **Flip Horizontal.**

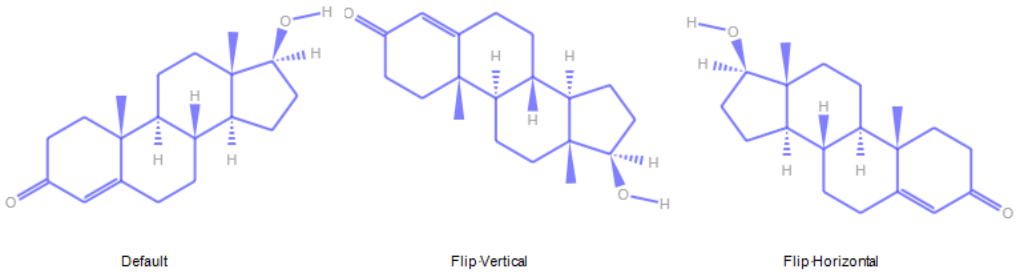


Figure 18. Flipping the diagram.

Move Atoms or Groups

The layout of the default diagram is sufficient for most purposes. However, it is sometimes useful to modify the layout—typically to improve readability—by moving

© 2011 Outercurve Foundation. All rights reserved.

atoms or groups of atoms. For example, when you replace an atom with a group of atoms, you can end up with a crowded or even overlapping diagram.

To modify the structure, select the atom or group of atoms that you want to move, and drag it to the new location. The editor will adjust the bonds accordingly. The left side of figure 19 shows the diagram after adding a COCl group.

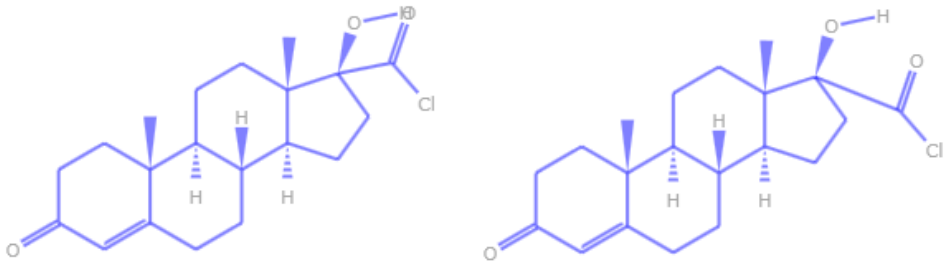


Figure 19. Moving atoms or groups

When you make this change, the oxygen atom ends up directly on top of a hydrogen atom. One way to make the diagram more readable is to move COCl group. The situation is complicated somewhat by the overlapping O and H atoms, which makes it difficult to select just the COCl group. The following procedure shows one way to address the problem and the resulting diagram is on the right side of Figure 19.

To move Cl and O

1. Select the COCl group by gesturing along the bond connecting the group to the ring.

   Using a gesture enables you to select the COCl group, without including the H atom.

2. Drag the COCl group to a better location.

Rotate Atoms or Groups

Moving atoms or groups, as shown in the previous section, simply translates the atoms. This can sometimes leave you with an awkward looking diagram. Another way to improve readability is to rotate the group, perhaps combined with translation.

To rotate a group

1. Select the group.
2. Click **Rotate** on the popup.
3. Drag the green button to rotate the group to a suitable orientation.

The right side of Figure 20 shows the same initial diagram as Figure 19, after selecting the COCl group and clicking Rotate. The right side shows the diagram after the COCl group has been rotated clockwise and translated slightly.

© 2011 Outercurve Foundation. All rights reserved.

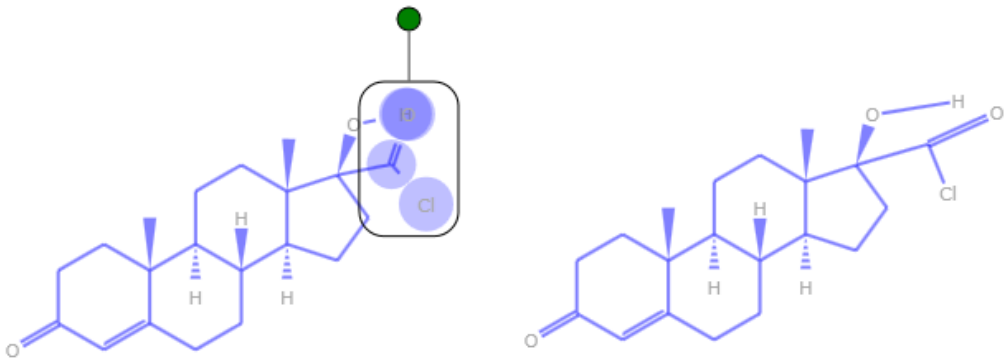


Figure 20. Rotating groups

Delete Atoms or Groups

In addition to the Delete command on the popup, the editor also includes a **Delete** button, on the left side of the window. You can use this button to quickly delete atoms or groups of atoms.

© 2011 Outercurve Foundation. All rights reserved.

Resources

This section provides links to additional information about the Chemistry Add-in.

2007 Microsoft Office system requirements
http://office.microsoft.com/en-us/products/HA101668651033.aspx

2010 Microsoft Office system requirements
http://technet.microsoft.com/en-us/library/ee624351(office.14).aspx

cml.sourceforge.net - OpenSource Site for CML
http://cml.sourceforge.net/

Ecma Office Open XML File Formats overview
http://office.microsoft.com/en-us/products/HA102058151033.aspx

ISO Standard
http://www.iso.org/iso/iso_catalogue.htm

Open XML Formats
http://www.openxmldeveloper.org

JUMBO FAQ
http://www.ch.ic.ac.uk/omf/cml/doc/jumbo/faq.html

Office Development with Visual Studio Developer Center
http://msdn.microsoft.com/en-us/vsto/default.aspx

PubChem
http://pubchem.ncbi.nlm.nih.gov/

© 2011 Outercurve Foundation. All rights reserved.

Appendix 1: CML Basics

CML is an XML-based format for representing a wide range of chemical data. The Chemistry Add-in uses a subset of CML, the molecular convention, which includes support for:

- IUPAC and other names
- Chemical structure
- Formulae

The following example shows the CML data for water.

```
<?xml version="1.0" ?>
<cml convention="conventions:molecular"
     xmlns="http://www.xml-cml.org/schema"
     xmlns:conventions="http://www.xml-cml.org/convention/"
     xmlns:nameDict="http://www.xml-cml.org/dictionary/cml/name/">
  <molecule id="m1">
    <name dictRef="nameDict:trivial">H_{2}O</name>
    <name dictRef="nameDict:iupac">oxidane</name>
    <name dictRef="nameDict:trivial">aqua</name>
    <name dictRef="nameDict:trivial">water</name>
    <name dictRef="nameDict:systematic">dihydrogen monoxide</name>
    <name dictRef="nameDict:systematic">hydrogen hydroxide</name>
    <name dictRef="nameDict:trivial">Adam's ale</name>
    <formula concise="H 2 O 1" />
    <atomArray>
      <atom id="a1" elementType="O"
            x2="-1.5950000286102295" y2="1.1549999713897705" />
      <atom id="a2" elementType="H"
            x2="-0.05500002861022946" y2="1.1549999713897705" />
      <atom id="a3" elementType="H"
            x2="-2.3650000286102295" y2="2.4886790932178062" />
    </atomArray>
    <bondArray>
      <bond id="b1" atomRefs2="a1 a2" order="1" />
      <bond id="b2" atomRefs2="a1 a3" order="1" />
    </bondArray>
  </molecule>
</cml>
```

The data is typical for the Chemistry Add-in molecules, and includes:

- The IUPAC name
- Several trivial and systematic names
- 2-D structural data, including atom positions and bonding details

For a summary of the commonly used CML elements in the molecular convention, see http://www.xml-cml.org/convention/. For details on CML, including the current schema and related tools, see http://www.xml-cml.org/ and “cml.sourceforge.net - OpenSource Site for CML.”

CML Data Storage

The Chemistry Add-in stores CML data in the following ways

© 2011 Outercurve Foundation. All rights reserved.

- The Chemistry Add-in Smart Tag folder contains CML data for several hundred common molecules.

To simplify chemistry zone creation, many molecules are represented by smart tags, as discussed in the next section. The associated data for each molecule is in a .cml file in the add-in's Smart Tag folder, which is typically c:\Program Files\Chem4Word\Smart Tag.

- The Chemistry Gallery is a Word gallery that contains a collection of chemistry zones.

The gallery typically enables users to quickly find and insert a zone into a document. It usually contains a small collection of commonly used molecules and users can customize the collection, as appropriate. The Chemistry Gallery has its own backing data; a set of .cml files in the add-in's c:\Program Files\Chem4Word\data folder.

- When a user inserts a chemistry zone into a document, the associated CML data is stored in the document.

Some commands modify the embedded CML data in various ways, such as adding trivial names or modifying the 2-D structure. However, these commands affect only the data that is stored in the document.

Appendix 2: How CML Data is Stored in a Document

A Word document is basically a structured collection of files, stored a ZIP-compatible Open Package Convention (OPC) format. The document has a .docx extension, but you can use Windows Explorer to examine the contents of a Word document by simply changing the extension to .zip.

Tip: Although you can see the contents of a .zip file in Windows Explorer, the contents are easier to work if you first extract them to a folder.

Figure 4 shows the contents of a a document named Chem4Word_Example, which was saved after creating a "benzene" chemistry zone.

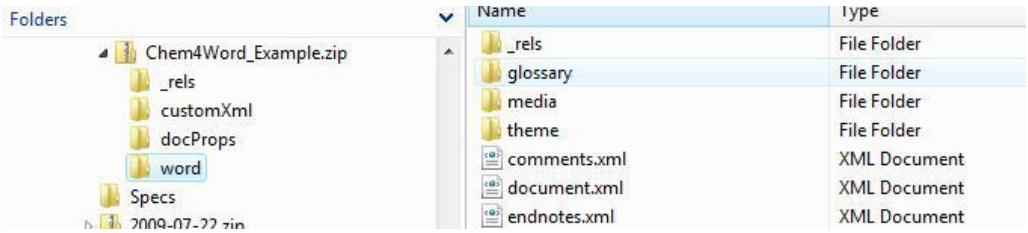


Figure 4. Word document structure

The document is primarily a collection of XML files. Figure 3 shows the Word folder, which contains the document's primary files. In particular, document.xml contains the body of the document in Office Open XML (OOXML) format. For more information on OOXML, see "Ecma Office Open XML File Formats overview."

Chemistry zones are represented in document.xml by a structured document element (<w:sdt>). The Chemistry Add-in identifies the associated XML block as a chemistry

© 2011 Outercurve Foundation. All rights reserved.

zone by setting the **alias** element's **val** attribute to “chemistry”. The following example shows the chemistry zone XML for benzene:

```
<w:sdt>
  <w:sdtPr>
    <w:alias w:val="chemistry"/>
    <w:id w:val="23589811"/>
    <w:placeholder>
      <w:docPart w:val="DefaultPlaceholder_22675703"/>
    </w:placeholder>
  </w:sdtPr>
  <w:sdtContent>
    <m:oMath>
      <m:r>
        <m:rPr>
          <m:sty m:val="p"/>
        </m:rPr>
        <w:rPr>
          <w:rFonts w:ascii="MS ChemSans" w:hAnsi="MS ChemSans "/>
        </w:rPr>
        <m:t>benzene</m:t>
      </m:r>
    </m:oMath>
  </w:sdtContent>
</w:sdt>
```

The zone’s CML data is stored separately, in the document’s customXml folder. The zone’s **id** element links the zone to the associated CML data file. Figure 5 shows a portion of the Chem4Word_Example’s customXml folder. The itemN.xml files contain various types of custom XML data, including the CML data. The numbering scheme is not related to the contents of the file. For this example, the benzene CML data is in item8.xml.

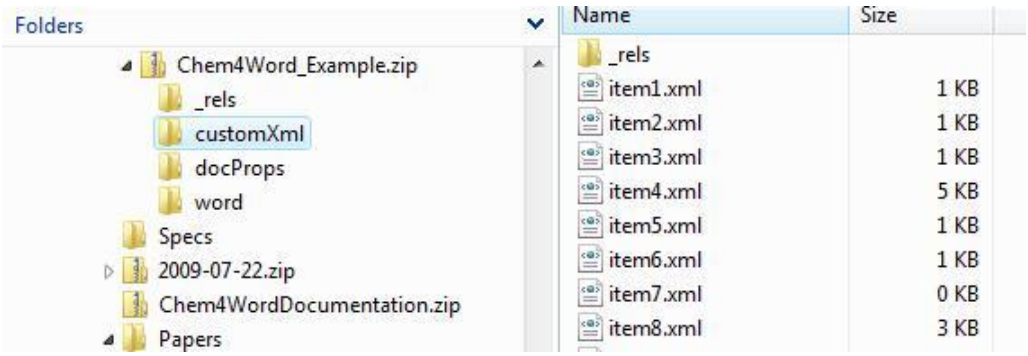


Figure 5. The customXml folder

The benzene chemistry zone was created with a smart tag, so item8.xml contains the CML-formatted data from benzene’s CML file in the Smart Tag folder, 241.cml.

```
<cml xmlns="http://www.xml-cml.org/schema"
 xmlns:nameDict="http://www.xml-cml.org/dictionary/cml/name/"
 xmlns:cmlDict="http://www.xml-cml.org/dictionary/cml/"
 xmlns:conventions="http://www.xml-cml.org/convention/"
 convention="conventions:molecular">
  <molecule id="m1" formalCharge="0" spinMultiplicity="1">
    <name dictRef="nameDict:iupac">benzene</name>
    <name dictRef="nameDict:trivial">Phenyl hydride</name>
    <formula concise="C 6 H 6"/>
```


© 2011 Outercurve Foundation. All rights reserved.

```
    <atomArray>
      <atom id="a1" elementType="C" x2="-1.924" y2="-5.774"/>
      <atom id="a2" elementType="C" x2="-3.258" y2="-6.544"/>
      <atom id="a3" elementType="C" x2="-3.258" y2="-8.085"/>
      <atom id="a4" elementType="C" x2="-1.924" y2="-8.855"/>
      <atom id="a5" elementType="C" x2="-0.591" y2="-8.085"/>
      <atom id="a6" elementType="C" x2="-0.591" y2="-6.544"/>
      <atom id="a7" elementType="H" x2="-1.924" y2="-4.234"/>
      <atom id="a8" elementType="H" x2="-4.592" y2="-5.774"/>
      <atom id="a9" elementType="H" x2="-4.592" y2="-8.855"/>
      <atom id="a10" elementType="H" x2="-1.924" y2="-10.395"/>
      <atom id="a11" elementType="H" x2="0.742" y2="-8.855"/>
      <atom id="a12" elementType="H" x2="0.742" y2="-5.774"/>
    </atomArray>
    <bondArray>
      <bond id="b1" atomRefs2="a1 a2" order="S" />
      <bond id="b2" atomRefs2="a2 a3" order="D" />
      <bond id="b3" atomRefs2="a3 a4" order="S" />
      <bond id="b4" atomRefs2="a4 a5" order="D" />
      <bond id="b5" atomRefs2="a5 a6" order="S" />
      <bond id="b6" atomRefs2="a1 a6" order="D" />
      <bond id="b7" atomRefs2="a1 a7" order="S" />
      <bond id="b8" atomRefs2="a2 a8" order="S" />
      <bond id="b9" atomRefs2="a3 a9" order="S" />
      <bond id="b10" atomRefs2="a4 a10" order="S"/>
      <bond id="b11" atomRefs2="a5 a11" order="S"/>
      <bond id="b12" atomRefs2="a6 a12" order="S"/>
    </bondArray>
  </molecule>
</cml>
```

This data set contains the following benzene-related data:

- The IUPAC name (benzene).
- A trivial name (phenyl hydride).
- The concise formula (C_6H_6).
- Data that can be used to create a 2-D structural diagram.

© 2011 Outercurve Foundation. All rights reserved.